Kriging

Allan Aasbjerg Nielsen
Informatics and Mathematical Modelling
Technical University of Denmark, Building 321
DK-2800 Lyngby, Denmark
http://www.imm.dtu.dk/∼aa
aa@imm.dtu.dk

16 August 2004

"Kriging" (after the South African mining engineer and Professor Danie Krige) is a term used for a family of methods for minimum error variance estimation. Consider a linear (or rather affine) estimate $\hat{z}_0 = \hat{z}(\boldsymbol{r}_0)$ at a location $\boldsymbol{r}_0$ based on N measurements $\boldsymbol{z} = [z(\boldsymbol{r}_1), \ldots, z(\boldsymbol{r}_N)]^T = [z_1, \ldots, z_N]^T$ at locations $\boldsymbol{r}_i$

$$\hat{z}_0 = w_0 + \sum_{i=1}^{N} w_i z_i = w_0 + \boldsymbol{w}^T \boldsymbol{z}, \tag{1}$$

where w_0 is a constant and w_i is the weight applied to z_i.

We consider z_i as particular realisations of random variables Z_i, $\boldsymbol{Z} = [Z(\boldsymbol{r}_1), \ldots, Z(\boldsymbol{r}_N)]^T = [Z_1, \ldots, Z_N]^T$. We think of $Z(\boldsymbol{r})$ as consisting of a deterministic mean and a stochastic residual $Z(\boldsymbol{r}) = \mu(\boldsymbol{r}) + \epsilon(\boldsymbol{r})$ with a constant covariance $C(\boldsymbol{r}, \boldsymbol{h}) = C(\boldsymbol{h})$ where $\boldsymbol{h} = \boldsymbol{r}_i - \boldsymbol{r}_j$ is a displacement vector measuring distance and direction between locations $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$. For the linear (affine) estimator we get

$$\hat{Z}_0 = w_0 + \boldsymbol{w}^T \boldsymbol{Z}. \tag{2}$$

The actual error $z_0 - \hat{z}_0$ is unknown but for the expectation of the estimation error we get

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \mathrm{E}\{Z_0 - w_0 - \boldsymbol{w}^T \boldsymbol{Z}\} = \mu_0 - w_0 - \boldsymbol{w}^T \boldsymbol{\mu}, \tag{3}$$

where $\mu_0 = \mu(\boldsymbol{r}_0)$ is the expectation value of Z_0 and $\boldsymbol{\mu}$ is a vector of expectation values

$$\boldsymbol{\mu} = \begin{bmatrix} \mu(\boldsymbol{r}_1) \\ \vdots \\ \mu(\boldsymbol{r}_N) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}. \tag{4}$$

We want our estimator to be unbiased, i.e. we request $\mathrm{E}\{Z_0 - \hat{Z}_0\} = 0$ or

$$\mu_0 - w_0 - \boldsymbol{w}^T \boldsymbol{\mu} = 0. \tag{5}$$

The estimation variance is

$$\sigma_{\mathrm{E}}^2 = \mathrm{Var}\{Z_0 - \hat{Z}_0\} = \mathrm{Var}\{Z_0\} + \mathrm{Var}\{w_0 + \boldsymbol{w}^T\boldsymbol{Z}\} - 2\mathrm{Cov}\{Z_0, w_0 + \boldsymbol{w}^T\boldsymbol{Z}\}$$
$$= \sigma^2 + \boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w} - 2\boldsymbol{w}^T\mathrm{Cov}\{\boldsymbol{Z}, Z_0\}, \tag{6}$$

where $\boldsymbol{C}$ is the variance/covariance matrix (also known as the dispersion matrix) of $\boldsymbol{Z}$, $\boldsymbol{C} = \mathrm{D}\{\boldsymbol{Z}\}$ with elements $C_{ij} = C(\boldsymbol{h}_{ij}) = C(\boldsymbol{r}_i - \boldsymbol{r}_j)$ that are the covariances between observations at locations $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$. $\mathrm{Cov}\{\boldsymbol{Z}, Z_0\}$ is a column vector of covariances between observations at locations $\boldsymbol{r}_i$ and $\boldsymbol{r}_0$. σ_{E}^2 is a quadratic function in $\boldsymbol{w}$.

Simple Kriging (SK)

In "simple kriging" we assume that $\mu(\boldsymbol{r})$ is known (often assumed constant). If we insert w_0 from (5) into (3) we get

$$\hat{Z}_0 - \mu_0 = \boldsymbol{w}^T(\boldsymbol{Z} - \boldsymbol{\mu}) \tag{7}$$

The kriging weights w_i are found by minimising σ_{E}^2

$$\frac{\partial \sigma_{\mathrm{E}}^2}{\partial \boldsymbol{w}} = 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{\boldsymbol{Z}, Z_0\} = \boldsymbol{0}, \tag{8}$$

which gives the simple kriging system

$$\boldsymbol{C}\boldsymbol{w} = \mathrm{Cov}\{\boldsymbol{Z}, Z_0\} \tag{9}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} \\ \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \end{bmatrix}. \tag{10}$$

The minimum squared estimation error is

$$\sigma_{\mathrm{SK}}^2 = \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{\boldsymbol{Z}, Z_0\}) = \sigma^2 - \boldsymbol{w}^T\mathrm{Cov}\{\boldsymbol{Z}, Z_0\}. \tag{11}$$

In SK the mean is known. In practice one must estimate $\mu(\boldsymbol{r})$ prior to the estimation or design an estimation algorithm that requires no prior mean.

Ordinary Kriging (OK)

In "ordinary kriging" we assume that the mean is constant ($= \mu_0$) for Z_0 and the N points that enter into the estimation of Z_0

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \mu_0(1 - \boldsymbol{w}^T\boldsymbol{1}) - w_0 = 0 \tag{12}$$

for any μ_0. $\boldsymbol{1}$ is a vector of ones. This is possible only if $w_0 = 0$ and $\boldsymbol{w}^T\boldsymbol{1} = 1$.

The weights w_i are found by minimising σ_{E}^2 with $\boldsymbol{w}^T\boldsymbol{1} = 1$. Introduce F with Lagrange multiplier -2λ

$$F = \sigma_{\mathrm{E}}^2 + 2\lambda(\boldsymbol{w}^T \mathbf{1} - 1) \tag{13}$$

$$\frac{\partial F}{\partial \boldsymbol{w}} = 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{\boldsymbol{Z}, Z_0\} + 2\lambda\mathbf{1} = \mathbf{0} \tag{14}$$

$$\frac{\partial F}{\partial \lambda} = 2(\boldsymbol{w}^T \mathbf{1} - 1) = 0 \tag{15}$$

which gives the ordinary kriging system

$$\boldsymbol{C}\boldsymbol{w} + \lambda\mathbf{1} = \mathrm{Cov}\{\boldsymbol{Z}, Z_0\} \tag{16}$$

$$\mathbf{1}^T\boldsymbol{w} = 1 \tag{17}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{N1} & \cdots & C_{NN} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \lambda \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \\ 1 \end{bmatrix}. \tag{18}$$

The minimum squared estimation error is

$$\sigma_{\mathrm{OK}}^2 = \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{\boldsymbol{Z}, Z_0\}) = \sigma^2 - \boldsymbol{w}^T\mathrm{Cov}\{\boldsymbol{Z}, Z_0\} - \lambda. \tag{19}$$

OK implies a re-estimation of μ_0 for each new estimation support.

Universal Kriging (UK)

In “universal kriging” we assume that the mean can be written as a linear combination of known functions (ideally determined by the physics of the problem dealt with)

$$\mu(\boldsymbol{r}) = \sum_{\ell=0}^{L} a_\ell f_\ell(\boldsymbol{r}) = \boldsymbol{a}^T \boldsymbol{f}(\boldsymbol{r}). \tag{20}$$

If no knowledge exists about the form of $\boldsymbol{f}$ we can use low order (typically second order) polynomials and consider them as local Taylor expansions. By convention $f_0 = 1$.

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \boldsymbol{a}^T\boldsymbol{f}(\boldsymbol{r}_0) - w_0 - \boldsymbol{w}^T\boldsymbol{\mu}. \tag{21}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu(\boldsymbol{r}_1) \\ \vdots \\ \mu(\boldsymbol{r}_N) \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}^T\boldsymbol{f}(\boldsymbol{r}_1) \\ \vdots \\ \boldsymbol{a}^T\boldsymbol{f}(\boldsymbol{r}_N) \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}^T(\boldsymbol{r}_1)\boldsymbol{a} \\ \vdots \\ \boldsymbol{f}^T(\boldsymbol{r}_N)\boldsymbol{a} \end{bmatrix} = \boldsymbol{H}\boldsymbol{a} \tag{22}$$

with

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{f}^T(\boldsymbol{r}_1) \\ \vdots \\ \boldsymbol{f}^T(\boldsymbol{r}_N) \end{bmatrix}, \quad \boldsymbol{H}^T = [\boldsymbol{f}(\boldsymbol{r}_1) \ \cdots \ \boldsymbol{f}(\boldsymbol{r}_N)]. \tag{23}$$

$$\mathrm{E}\{Z_0 - \hat{Z}_0\} = \boldsymbol{a}^T\boldsymbol{f}(\boldsymbol{r}_0) - w_0 - \boldsymbol{w}^T\boldsymbol{H}\boldsymbol{a} = \boldsymbol{a}^T(\boldsymbol{f}(\boldsymbol{r}_0) - \boldsymbol{H}^T\boldsymbol{w}) - w_0 = 0 \tag{24}$$

for any $\boldsymbol{a}$. This is possible only if $w_0 = 0$ and $\boldsymbol{H}^T\boldsymbol{w} = \boldsymbol{f}(\boldsymbol{r}_0)$.

The weights w_i are found by minimising σ_{E}^2 with $\boldsymbol{H}^T\boldsymbol{w} = \boldsymbol{f}_0$. Introduce F with $L+1$ Lagrange multipliers $-2\boldsymbol{\lambda}$ where $\boldsymbol{\lambda} = [\lambda_0, \ldots, \lambda_L]^T$

$$F = \sigma_{\mathrm{E}}^2 + 2\boldsymbol{\lambda}^T(\boldsymbol{H}^T\boldsymbol{w} - \boldsymbol{f}(\boldsymbol{r}_0)) \tag{25}$$

$$\begin{aligned} \frac{\partial F}{\partial \boldsymbol{w}} &= 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{\boldsymbol{Z}, Z_0\} + 2\boldsymbol{H}\boldsymbol{\lambda} = \boldsymbol{0} && (26) \\ \frac{\partial F}{\partial \boldsymbol{\lambda}} &= 2(\boldsymbol{H}^T\boldsymbol{w} - \boldsymbol{f}(\boldsymbol{r}_0)) = \boldsymbol{0} && (27) \end{aligned}$$

which gives the universal kriging system

$$\begin{aligned} \boldsymbol{C}\boldsymbol{w} + \boldsymbol{H}\boldsymbol{\lambda} &= \mathrm{Cov}\{\boldsymbol{Z}, Z_0\} && (28) \\ \boldsymbol{H}^T\boldsymbol{w} &= \boldsymbol{f}(\boldsymbol{r}_0) && (29) \end{aligned}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} & f_0(\boldsymbol{r}_1) & \cdots & f_L(\boldsymbol{r}_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{NN} & f_0(\boldsymbol{r}_N) & \cdots & f_L(\boldsymbol{r}_N) \\ f_0(\boldsymbol{r}_1) & \cdots & f_0(\boldsymbol{r}_N) & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_L(\boldsymbol{r}_1) & \cdots & f_L(\boldsymbol{r}_N) & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \lambda_0 \\ \vdots \\ \lambda_L \end{bmatrix} = \begin{bmatrix} C_{01} \\ \vdots \\ C_{0N} \\ f_0(\boldsymbol{r}_0) \\ \vdots \\ f_L(\boldsymbol{r}_0) \end{bmatrix}. \tag{30}$$

The minimum squared estimation error is

$$\sigma_{\mathrm{UK}}^2 = \sigma^2 + \boldsymbol{w}^T(\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{\boldsymbol{Z}, Z_0\}) = \sigma^2 - \boldsymbol{w}^T\mathrm{Cov}\{\boldsymbol{Z}, Z_0\} - \boldsymbol{\lambda}^T\boldsymbol{f}(\boldsymbol{r}_0). \tag{31}$$

When applying UK $\boldsymbol{C}$ and $\mathrm{Cov}\{\boldsymbol{Z}, Z_0\}$ must be estimated after removal of the trend that we are estimating along with the stochastic part. This is a problem that limits the use of UK (one could argue that this characteristic doesn't go too well with the adjective "universal").

We can use UK to apply an external drift for example by setting $L = 1$

$$\boldsymbol{a}^T\boldsymbol{f}(\boldsymbol{r}) = \sum_{\ell=0}^{L} a_\ell f_\ell(\boldsymbol{r}) = a_0 + a_1 f_1(\boldsymbol{r}). \tag{32}$$

An external drift is easily allowed for before estimating $\boldsymbol{C}$ and $\mathrm{Cov}\{\boldsymbol{Z}, Z_0\}$.

Factorial Kriging (FK)

In "factorial kriging" we assume that the random variable can be written not as a sum of a deterministic mean and a stochastic error but rather as a sum of $K+1$ independent stochastic terms or "factors" Y_i

$$Z(\boldsymbol{r}_0) = Y_0(\boldsymbol{r}_0) + Y_1(\boldsymbol{r}_0) + \cdots + Y_K(\boldsymbol{r}_0). \tag{33}$$

Since the factors are independent we get for the covariance function for Z

$$C(\boldsymbol{h}) = C_0(\boldsymbol{h}) + C_1(\boldsymbol{h}) + \cdots + C_K(\boldsymbol{h}) \tag{34}$$

where $C_i(\boldsymbol{h})$ is the covariance function for Y_i. The factors are ordered so that low indices correspond to short range phenomena and high indices correspond to phenomena with increasingly longer range for example stemming from $K+1$ nested structures identified in the model of $C(\boldsymbol{h})$.

We can filter out the covariance contributions of any number of consecutive short range or noise factors (with indices ranging from 0 to $k_0 - 1$ with $0 < k_0 < K$) leaving the long range or signal factors only.

The long range or signal component of Z is $Z_{\mathrm{LR}} = \sum_{k=k_0}^{K} Y_i$. Its estimator is $\hat{Z}_{\mathrm{LR}} = \boldsymbol{w}^T \boldsymbol{Z}$. We want the expectation $\mathrm{E}\{Z_{\mathrm{LR}} - \hat{Z}_{\mathrm{LR}}\}$ to be zero. As in the OK case this leads to $\boldsymbol{w}^T \boldsymbol{1} = 1$.

The estimation variance in this case is

$$\sigma_{\mathrm{E}}^2 = \mathrm{Var}\{Z_{\mathrm{LR}} - \hat{Z}_{\mathrm{LR}}\} = \mathrm{Var}\{Z_{\mathrm{LR}}\} + \boldsymbol{w}^T \boldsymbol{C} \boldsymbol{w} - 2\boldsymbol{w}^T \mathrm{Cov}\{\boldsymbol{Z}, Z_{\mathrm{LR}}\}. \tag{35}$$

The weights w_i are found by minimising σ_{E}^2 with $\boldsymbol{w}^T \boldsymbol{1} = 1$. Introduce F with Lagrange multiplier -2λ

$$F = \sigma_{\mathrm{E}}^2 + 2\lambda(\boldsymbol{w}^T \boldsymbol{1} - 1) \tag{36}$$

$$\begin{aligned} \frac{\partial F}{\partial \boldsymbol{w}} &= 2\boldsymbol{C}\boldsymbol{w} - 2\mathrm{Cov}\{\boldsymbol{Z}, Z_{\mathrm{LR}}\} + 2\lambda \boldsymbol{1} = \boldsymbol{0} && (37) \\ \frac{\partial F}{\partial \lambda} &= 2(\boldsymbol{w}^T \boldsymbol{1} - 1) = 0 && (38) \end{aligned}$$

which gives the factorial kriging system (for the long range part)

$$\begin{aligned} \boldsymbol{C}\boldsymbol{w} + \lambda \boldsymbol{1} &= \mathrm{Cov}\{\boldsymbol{Z}, Z_{\mathrm{LR}}\} && (39) \\ \boldsymbol{1}^T \boldsymbol{w} &= 1 && (40) \end{aligned}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{N1} & \cdots & C_{NN} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \lambda \end{bmatrix} = \begin{bmatrix} C_{01\mathrm{LR}} \\ \vdots \\ C_{0N\mathrm{LR}} \\ 1 \end{bmatrix}. \tag{41}$$

The elements in the RHS vector are $C_{0i\mathrm{LR}} = \sum_{k=k_0}^{K} C_k(\boldsymbol{h}_{0i}),\ \boldsymbol{h}_{0i} = \boldsymbol{r}_0 - \boldsymbol{r}_i$.

$Z_{\text{SR}} = \sum_{k=0}^{k_0-1} Y_i$ is the short range or noise part of Z. Its estimator is $\hat{Z}_{\text{SR}} = \boldsymbol{w}^T \boldsymbol{Z}$. We want the expectation $\text{E}\{Z_{\text{SR}} - \hat{Z}_{\text{SR}}\}$ to be zero. Since we want the noise expectation $\text{E}\{Z_{\text{SR}}\}$ to be zero $\text{E}\{\hat{Z}_{\text{SR}}\}$ is zero also. This leads to $\boldsymbol{w}^T \mathbf{1} = 0$.

The estimation variance in this case is

$$\sigma_{\text{E}}^2 = \text{Var}\{Z_{\text{SR}} - \hat{Z}_{\text{SR}}\} = \text{Var}\{Z_{\text{SR}}\} + \boldsymbol{w}^T \boldsymbol{C} \boldsymbol{w} - 2\boldsymbol{w}^T \text{Cov}\{\boldsymbol{Z}, Z_{\text{SR}}\}. \tag{42}$$

The weights w_i are found by minimising σ_{E}^2 with $\boldsymbol{w}^T \mathbf{1} = 0$. Introduce F with Lagrange multiplier -2λ

$$F = \sigma_{\text{E}}^2 + 2\lambda \boldsymbol{w}^T \mathbf{1} \tag{43}$$

$$\frac{\partial F}{\partial \boldsymbol{w}} = 2\boldsymbol{C}\boldsymbol{w} - 2\text{Cov}\{\boldsymbol{Z}, Z_{\text{SR}}\} + 2\lambda \mathbf{1} = \mathbf{0} \tag{44}$$

$$\frac{\partial F}{\partial \lambda} = 2\boldsymbol{w}^T \mathbf{1} = 0 \tag{45}$$

which gives the factorial kriging system (for the short range part)

$$\boldsymbol{C}\boldsymbol{w} + \lambda \mathbf{1} = \text{Cov}\{\boldsymbol{Z}, Z_{\text{SR}}\} \tag{46}$$

$$\mathbf{1}^T \boldsymbol{w} = 0 \tag{47}$$

or

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{N1} & \cdots & C_{NN} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \lambda \end{bmatrix} = \begin{bmatrix} C_{01\text{SR}} \\ \vdots \\ C_{0N\text{SR}} \\ 0 \end{bmatrix}. \tag{48}$$

The elements in the RHS vector are $C_{0i\text{SR}} = \sum_{k=0}^{k_0-1} C_k(\boldsymbol{h}_{0i}),\ \boldsymbol{h}_{0i} = \boldsymbol{r}_0 - \boldsymbol{r}_i$.

We see that compared to the OK system, in the FK system the LHS is unchanged and that the covariances on the RHS are replaced with relevant long or short range components, and that for the long range part the weights sum to one as in OK and for the short range part they sum to zero.

References

A.G. Journel and Ch.J. Huijbregts (1978). *Mining Geostatistics.* Academic Press.

E.H. Isaaks and R.M. Srivastava (1989). *Applied Geostatistics.* Oxford University Press.

C.V. Deutsch and A.G. Journel (1998). *GSLIB: Geostatistical Software Library and User's Guide*, second edition. Oxford University Press.

T. Yao, T. Mukerji, A. Journel and G. Mavko (1999). Scale matching with factorial kriging for improved porosity estimation from seismic data. *Mathematical Geology* **31**(1), 23-46.